Prediction of 2024 US election ...*

Colin Sihan Yang Lexun Yu Siddharth Gowda

November 3, 2024

We forecast the winner of the 2024 US presidential election using "poll-of-polls" by building a linear model.

1 Introduction

Election result forecasting has become an essential tool for analysts in political science and the public to predict the outcome of democratic process, such as the presidential election in the United States. Traditionally, individual polls have been used as a snapshot of voter sentiment, but they only reflect temporary changes in the performance of contestants, instead of a precise estimation of the election result. As discussed by Pasek (2015) and Blumenthal (2014), the aggregation of multiple polls, or "poll-of-polls," has become a popular technique to reduce individual survey errors and provide more accurate election forecasts. However, the traditional poll aggregation does not reflect dynamics of an election, especially with real-time changes and the introduction of new data. This creates a gap for a more adaptable model to predict the election result based on both polling data and additional variables, such as historical data and economic indicators.

This paper fills the gap by building a hybrid election forecasting model following the strategies mentioned by Pasek (2015). As Pasek (2015) described in their article, aggregation involves determining which surveys are worth including, as well as selecting, combining and averaging results from multiple polls to reduce individual biases and errors. Prediction modeling adds other data to the model that predicts election outcomes based on current dynamics. Hybrid models like the Bayesian approach incorporates prior beliefs based on historical data or expert knowledge and new evidence like economic updates to dynamically adjust the forecast as the campaign progresses.

In this paper, we aim to predict the 2024 us election result with the hybrid election forcasting model. We incorporate aggregation by filtering the polls on FiveThirtyEight (2024) by

*Code and data are available at: https://github.com/yulexun/uselection.

numeric grade that indicates pollster's reliability, prediction that incorporates social and economic indicators including unemployment rates and abortion rates, and hybrid approaches that leverages Bayesian techniques which combines historical data such as the 2016 election data, allowing for a dynamic prediction of the U.S. presidential election.

The estimand for this research paper is the predicted support percentages for Kamala Harris and Donald Trump. The prediction is based on quantifying various polling factors, including sample size, poll scores, and transparency scores, which are used as predictors.

The results of this model indicate a more stable and accurate forecast compared to traditional aggregation methods alone, [update this …]

The remainder of this paper is structured as follows: [update this …]

Appendix

2 Additional data details

3 Model details

3.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows…

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows…

3.2 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows… This suggests…

?@fig-stanareyouokay-2 is a Rhat plot. It shows… This suggests…

4 FiveThirtyEight Licenses

5 Our Prediction

Our prediction process consists of two primary components. First, we develop models for both Trump and Harris based on the variables outlined in Section **?@sec-PLACEHOLDER**. This involves partitioning the dataset into training and testing subsets. Next, we further divide the testing dataset into swing states and other competitive races. We then input this test data into the respective models to generate predictions. By averaging these predictions, we can calculate the expected voter percentage for each candidate in each state. The candidate with the higher percentage is deemed the winner for that state.

We generated predictions for the following states: Arizona, Nevada, Georgia, Pennsylvania, Michigan, Minnesota, Wisconsin, Florida, Texas, Maine CD-2, Nebraska CD-2, New Hampshire, Ohio, Virginia, North Carolina, and Iowa. Winners for other states were determined based on historical trends and predictions from sources like (**cnn?**). Most states without predictions are strongly Republican or Democratic, so their absence is not expected to significantly impact prediction validity.

State	Harris Predicted Percentage	Trump Predicted Percentage	State Winner
Arizona	46.59088	49.22984	Trump
Florida	42.86648	50.75718	Trump
Georgia	47.23656	48.88107	Trump
Iowa	48.55001	43.55272	Harris
Maine CD-2	47.40416	49.31394	Trump
Michigan	47.59588	46.98425	Harris
Minnesota	48.59180	43.69296	Harris
Nebraska CD-2	49.90928	42.17853	Harris
Nevada	49.38011	47.12200	Harris
New Hampshire	50.77677	42.64558	Harris
North Carolina	48.70390	47.72535	Harris
Ohio	43.95410	50.98532	Trump
Pennsylvania	48.22853	47.15005	Harris
Texas	44.98253	50.22253	Trump
Virginia	49.16028	43.20352	Harris
Wisconsin	48.42638	46.44137	Harris

According to Figure Figure 1, Kamala Harris is predicted to be the 47th President of the United States. There are also a few states with predictions not visible in the map, we will describe those predicts them below

In Maine, Harris is projected to win the state's overall delegates and District 1, while District 2 is expected to go to Trump. In Nebraska, Trump is expected to win the state's delegates

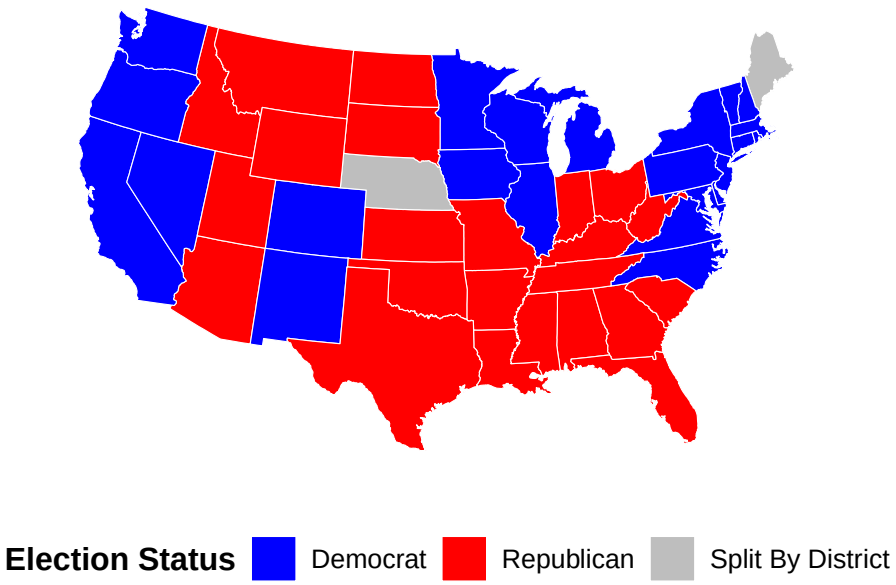


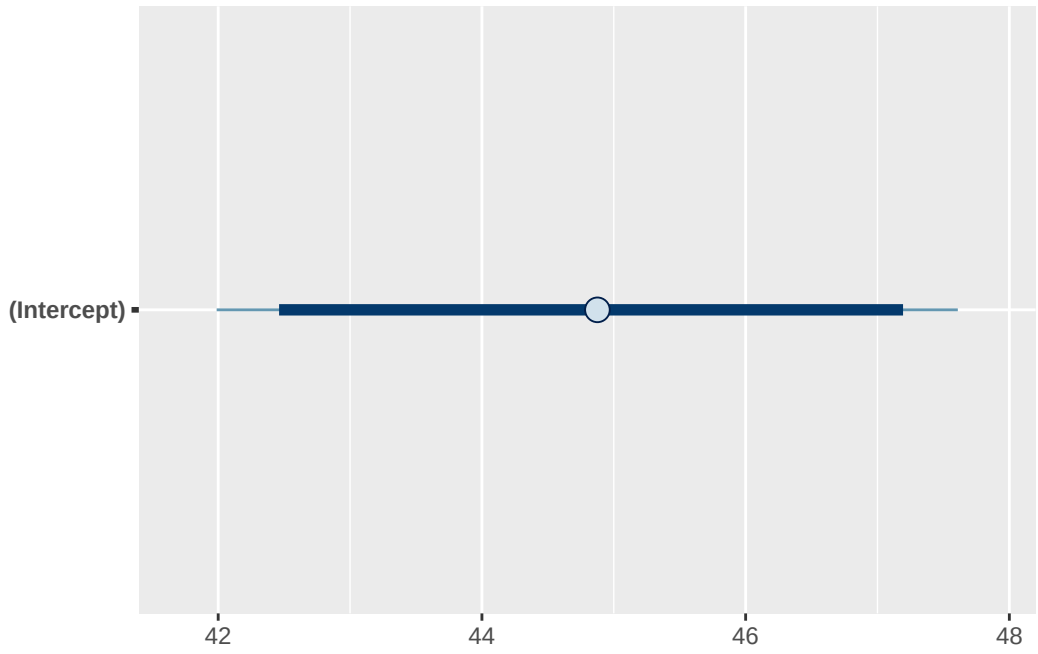
Figure 1: Kamala Harris is Predicited to be the 47th President of the United States

along with Districts 1 and 3, while Harris is predicted to win District 2. Additionally, Trump is projected to win Alaska, and Harris is expected to win Hawaii.

Overall, the predictions indicate that Harris will receive 298 delegates, while Trump will receive 240 delegates.

```
plot(bayesian_model_train_trump, pars = "(Intercept)", prob = 0.95)
```

```
Warning: `prob_outer` (0.9) is less than `prob` (0.95)
... Swapping the values of `prob_outer` and `prob`
```


(Intercept)
42
44
46
48

References

Blumenthal, Mark. 2014. "Polls, Forecasts, and Aggregators." *PS: Political Science and Politics* 47 (2): 297–300. http://www.jstor.org/stable/43284537.

FiveThirtyEight. 2024. "Our Data." *FiveThirtyEight*. https://data.fivethirtyeight.com.

Pasek, Josh. 2015. "THE POLLS–REVIEW: PREDICTING ELECTIONS: CONSIDERING TOOLS TO POOL THE POLLS." *The Public Opinion Quarterly* 79 (2): 594–619. http://www.jstor.org/stable/24546379.